5

ADAPTIVE SAMPLING TECHNIQUE FOR SELECTING NEGATIVE EXAMPLES FOR ARTIFICIAL INTELLIGENCE APPLICATIONS

A PATENT APPLICATION BY:

10   INVENTORS:

SRINIVAS GUTTA

K. P. LEE

J. DAVID SCHAFFER

15   DRAFTING ATTORNEY

ANNE E. BARSCHALL

REG. NO. 31,089

I. BACKGROUND OF THE INVENTION

A. Field of the Invention

The invention relates to artificial intelligence applications that require training sets having

positive and negative examples, especially recommender systems and particularly for use with

5   television. The invention relates even more particularly to such applications that use statistically

valid techniques for choosing negative examples for training sets.

B. Related Art

U.S. Patent Application Serial Number 09/498,271 filed 2/4/00 (US 000018),

10   incorporated herein by reference, discloses a television recommender system. In that system,

recommendations are deduced based on a pattern of shows watched or not watched. Of course,

the number of shows not watched necessarily dwarfs the number of shows watched.

Accordingly, a heuristic was developed for selecting shows not watched. The heuristic was to

select a not watched show for each watched show, the not watched show being taken at random

15   from time slots other than the slot in which the corresponding watched show occurred.

In general, many artificial intelligence applications have training sets with positive and

negative examples. The heuristic for selecting negative examples needs improvement over the

concept of selecting negative examples at random one-by-one with reference to respective

individual positive examples.

20   .

II. SUMMARY OF THE INVENTION

It is an object of the invention to improve heuristics for choosing negative examples for a training set for an artificial intelligence application.

This object is achieved in that a group of negative examples is selected corresponding to a group of positive examples, rather than one-by-one.

This object is further achieved in that the group of positive examples is analyzed according to a feature presumed to be dominant. Then a first fraction of the negative examples is taken from the non-positive possible examples sharing the feature with the positive examples.

This object is still further achieved in that a second fraction of the shows is taken from slots within a predetermined range in feature space with respect to the feature.

This object is yet still further achieved in that no negative example is taken more than once.

Advantageously the application in question is a recommender for content, such as television, where the positive examples are selected content, the negative examples are non-selected content. Advantageously, also, the feature is time of day of broadcast.

Further objects and advantages will be apparent in the following.

III. BRIEF DESCRIPTION OF THE DRAWING

The invention will now be described by way of non-limiting example with reference to the following drawings.

Fig. 1 graphs how negative examples are chosen using uniform random sampling, with

respect to a particular viewer, known as user H.

Fig. 2 shows a histogram of positive examples corresponding to Fig. 1.

Fig. 3 shows a histogram of negative examples corresponding to Fig. 1.

Fig. 4 is analogous to Fig. 1, except with respect to user C.

Fig. 5 is analogous to Fig. 2, except with respect to user C.

Fig. 6 is analogous to Fig. 3, except with respect to user C.

Fig. 7 is analogous to Fig. 1, except using the invention to select negative examples.

Fig. 8 is analogous to Fig. 4, except using the invention to select negative examples.

Fig. 9 is analogous to Fig. 3, except using the invention to select negative examples.

Fig. 10 is analogous to Fig. 6, except using the invention to select negative examples.

Fig. 11 shows the hit rate for user H as a function of false positive rate.

Fig. 12 is analogous to Fig. 11, except with respect to user C.

Fig. 13 shows hardware for implementing the invention.

Fig. 14 shows a flowchart of a process for building a training set in accordance with the invention.

Fig. 15 shows a table.

Fig. 16 shows a table

IV. DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Herein the invention is described with respect to recommenders for television, but it might be equally applicable to training sets for any artificial intelligence application, including

recommenders for other types of content. The term "show" is intended to include any other type

of content that might be recommended by a recommender, including audio, software, and text

information. The term "watch" or "watched" is intended to include any type of positive example

selection, including experiencing of any type of content, e.g. listening or reading. The invention

5   is described also with the assumption that time is the dominant feature distinguishing watched

from not-watched content; however other dominant features might be used as parameters for

selecting negative examples for a training set.

Fig 13 illustrates hardware for implementing the invention. The hardware will typically

have a display 1; some type of processor 2; at least one user entry device 4 connected to the

10   processor via some type of connection 3; and some type of link 5 for receiving data, such as

television programming or Electronic Programming Guide ("EPG") data. The display 1 will

commonly be a television screen, but could be any other type of display device. The processor 2

may be a set top box, a PC, or any other type of data processing device, so long as it has

sufficient processing power. The user entry device 4 may be a remote and the connection 3 may

15   be a wireless connection such as an infrared connection.   If the processor is a PC, the user entry

device will commonly be at least plural, e.g. a keyboard and a pointer device such as a mouse.

The user entry device may also be a touch sensitive display. The connection 5 to the outside

world could be an antenna, cable, a phone line to the internet, a network connection, or any other

data link. Moreover, the link 5 may allow communication with many different types of devices

20   such as remote processors, peripherals, and/or memory devices.

Commonly there will be at least one memory device 6, such as a CD ROM drive, floppy

disk drive, hard drive, or any other type of memory device. The memory device 6 can store data, software, or both.

There may be other peripherals not shown, such as a voice recognition system, a PC camera, speakers, and/or a printer.

5   Fig. 1 shows how negative examples are chosen using uniform random sampling, with respect to a particular viewer known as User H. The vertical axis shows time of day of the show. The horizontal axis shows the ordinal number of the samples. The circles are watched shows and the crosses are the corresponding unwatched samples. It can be seen that the watched shows are primarily clustered in the prime time part of the evening, with a smattering of shows watched

10   at other times of day, especially the very early morning, presumably before the viewer left for work or school.

Fig. 2 shows a histogram of positive examples, i.e. the cardinal number shows watched by User H, plotted against time of day, corresponding to the circle data illustrated in Fig. 1. With respect to Figures 2, 3, 5, 6, 9, and 10, it should be noted that the horizontal scale is only

15   approximate. The bars should not be considered as corresponding exactly to the times listed below them.

Fig. 3 shows a histogram of negative examples, i.e. not watched shows, using uniform random sampling. Again a cardinal number of shows is plotted against time of day. This data corresponds to the crosses of Fig. 1.

20   Fig. 4 is the same type of data as in Fig. 1, except that it is taken with respect to a second user, called User C. There are fewer samples taken for this user than for user H. User C has 175

samples, while User H had more than 275.

Fig. 5 is like Fig. 2, except with respect to User C.

Fig. 6 is like Fig. 3, except with respect to User C.

Fig. 14 shows a flowchart of a process for building a training set in accordance with the

5   invention. The operations of this flowchart may be executed on the processor 2, or any

processor connected with or receiving data from processor 2, e.g. via link 5. Similarly, the

artificial intelligence application itself, e.g. the content recommender, may be trained and run on

the processor 2 or any processor connected with or receiving data from processor 2, e.g. link 5.

At 1401, a population of watched shows of statistically significant size is accumulated. In

10   the examples of Users H & C, the sizes of the population are over 275 and 175, respectively;

however, other size populations are usable, so long as they are statistically significant.

At 1402 the distribution of watched shows with respect to time is determined, and

preferred time slots are determined. The distribution can be taken in the form of a histogram, see

e.g. Fig. 2 or Fig. 5. In the preferred embodiment, the five time slots having the most shows are

15   chosen. However, more or fewer preferred time slots may be chosen. Optionally, all the time

slots viewed by the user may be used. In the examples, the five most preferred time slot for User

H will be 21:00, 20:00, 19:00, 23:00, and 22:00, in that order, and the five most preferred time

slots for User C will be 8:00, 23:00, 20:00, 24:00, and 10:00, in that order.

Then at 1403, a first fraction of the negative examples is chosen in the preferred time

20   slots of this user. In the preferred embodiment the fraction is 50%.

At 1404, optionally, a second fraction of negative examples is taken from a

predetermined time interval around the preferred time slot or slots. In the preferred embodiment, the second fraction will be taken from the hour immediately before and the hour immediately after the single most preferred time slot If optional step 1404 is omitted, then all of the negative examples should be taken from either the preferred time slots, or from all the time slots viewed

5   by that user. Thus, the option -- at 1402 -- of using all the time slots used by the user is most likely to be chosen when step 1404 is to be omitted.

The negative example set is then taken at 1405 to include the first fraction and any second fraction. In the preferred embodiment, the negative example set in fact is just the first and second fractions.

10   At 1406, the recommender is trained using positive and negative example sets. Fig. 7 shows the same type of graph for user H as Fig. 1, except this time the negative examples are chosen in accordance with the technique of Fig. 14. It will be noted that the negative examples fall essentially where the positive examples fall in terms of time. The apparent monotonic curve in the negative examples is only an artifact of the order in which the negative examples are

15   chosen. They need not be chosen in any particular order.

Fig. 8 show the same type of graph for user C as Fig. 4, except this time the negative examples are chosen in accordance with the technique of Fig. 14. Again the apparent monotonic curve of the negative examples has no significance as it is only an artifact of the order in which the negative examples are chosen. Fig. 9 is like Fig. 3 with respect to User H, except that the

20   negative examples are chosen in accordance with the technique of Fig. 14. The histogram of positive examples for User H is not repeated here, because it is the same as before.

Fig. 10 is like Fig. 6 for User C, except that the negative examples are chosen in accordance with the technique of fig. 14.

The technique of Fig. 14 is shown experimentally to achieve an average performance increase of more than 20%. Performance is measured as the accuracy of recommendations for a set of TV shows for which users' ground truth assessments have been collected. Fig. 15 shows a table which is useful in understanding how performance is evaluated. The table defines four terms:

1. The system predicts a yes answer and there actually is a yes answer (TP)

2. The system predicts a no answer and there is actually a yes answer (FN)

3. The system predicts a yes answer and there actually is a no answer (FP)

4. The system predicts a no answer and there actually is a no answer (TN)

Then the "hit rate" is defined in accordance with the following equation:

$$Hit\ rate = \frac{TP}{TP + FP}$$

And the false positive rate is calculated in accordance with the following equation:

$$false\ positive\ rate = \frac{FP}{FP + TN}$$

Fig. 11 shows the hit rate for user H as a function of false positive rate. The stars show the curve for sampling in accordance with adaptive sampling, while the circles show the curve

for sampling where negative examples are taken in a uniform random distribution.

Usually a content recommender will first assign a probability of success for each piece of content, with respect to a user. Then the content will be recommended if its probability of success exceeds some threshold. The points on the curve of Fig. 11 correspond to different thresholds. Table 2 of Fig. 16 shows a further explanation of the type of calculation which leads to Fig. 11. In this table, values of numbers of hits, numbers of false negatives, numbers of true rejections, numbers of false positives, hit rate, and false positive rate are shown with respect to various values of the threshold, i.e. taken from 0 to 1 in steps of .05. The values of Table 2 are actually for User C, using negative examples chosen in accordance with the invention.

Fig. 12 shows the same curves with respect to User C. Both techniques work better for User H than for User C, because the population of positive examples was larger for User H than for User C; however, in both cases, the negative example set in accordance with adaptive sampling gives at least a 20% improvement.

In the examples given above, the set of negative examples is generally chosen to have the same number of members as the set of positive examples. However, those of ordinary skill in the art can design training sets in accordance with the invention where the number of negative examples is more or less than the number of positive examples.

From reading the present disclosure, other modifications will be apparent to persons skilled in the art. Such modifications may involve other features which are already known in the design, manufacture and use of training sets for artificial intelligence applications and which may

be used instead of or in addition to features already described herein. Although claims have been formulated in this application to particular combinations of features, it should be understood that the scope of the disclosure of the present application also includes any novel feature or novel combination of features disclosed herein either explicitly or implicitly or any generalization thereof, whether or not it mitigates any or all of the same technical problems as does the present invention. The applicants hereby give notice that new claims may be formulated, including method, software embodied in a storage medium, and "means for" claims, to such features during the prosecution of the present application or any further application derived therefrom.

The word "comprising", "comprise", or "comprises" as used herein should not be viewed as excluding additional elements. The singular article "a" or "an" as used herein should not be viewed as excluding a plurality of elements.